



## Guidelines for Summative Assignment

### Data analysis assignment: univariate and bivariate statistical tests and results interpretation

- ▼ 30% of final grade
- ▼ Due after week 10
  - ▼ **Deadline: 5<sup>th</sup> May, 2023**
- ▼ **Submission Opens: 24/04/2023 (Week 10)**

#### 1. General Information:

As part of this assignment you will be asked to identify and apply the appropriate statistical tests for assessing the association between two variables. Furthermore, you will be asked to interpret and draw conclusions from the results of these tests. Lastly, you will be asked to reflect on your process of learning in the course so far.

Each student will need to perform all of the following tasks in these five (5) questions, and provide their answers in a comprehensive report.

The structure of the report will follow the order of the tasks and specific guidelines, as they are outlined below. There is no word-limit.

The assignment covers sections 1-10 of the MPH-512 course.

An online webinar will be dedicated to introducing the assignment and helping students apply the course material in order to perform the required tasks and prepare their report.

## 2. Tasks and Specific Guidelines

You are provided with the STATA dataset titled:

*“MPH512\_example\_dataset\_smoke.dta”*

The dataset includes smoking related variables collected during a routine clinic visit for 230 individuals.

Use this dataset to perform the tasks described under the following 4 questions. Create and attach one Word file with your answers.

NOTE! Where you are asked to perform a task in STATA, please provide the relevant output as a screenshot or as copy-paste (use **Courier New** font to keep the tables aligned).

After you are finished with the exercise, save the dataset and clear Stata’s memory

Specific Instructions:

Open STATA and load the dataset *MPH512\_example\_dataset\_smoke.dta*

**Question 1**

Researchers wished to investigate the smoking status ("*smoker*"), as a categorical predictor variable, with respect to its association with being overweight or obese:

- a. Generate the variable "BMI\_category" and tabulate it. Under this variable:
  - Indicate all participants with *bmi* under 25.00 as "1"
  - Indicate all participants with *bmi* 25.00 and above, but less than 30.00 as "2"
  - Indicate all participants with *bmi* 30.00 and above as "3"
  - Indicate all missing values as "."

Provide a screenshot of the above commands and table.

- b. Create labels for the BMI\_category variable to indicate those with a 1 value as "normal", those with a 2 value as "overweight", and those with a 3 value as "obese". Attach the labels to the variable. Re-tabulate the variable BMI\_category.

Provide a screenshot of the label creation commands and the table.

- c. Run the most appropriate bivariate test to test this relationship (between smoking status and BMI category) and provide the relevant screenshot. Justify your choice of test.
- d. Interpret the results. Is there a significant association between smoking status and BMI category? If yes, interpret the direction of the relationship / effect estimate.

**Question 2**

Following, the study group wanted to investigate whether BMI category is associated with systolic blood pressure (“*sysbp\_before*”) in the study participants. Therefore, they wished to investigate whether systolic blood pressure differed significantly between BMI categories (normal vs. overweight vs. obese).

- a. Identify the independent/exposure variable and the dependent/outcome variable. State the type of each of these two variables
- b. Identify a parametric univariate test to test this relationship, justifying your choice of test in terms of the variables involved. Run the test and provide the relevant screenshot.
- c. Interpret the results of this parametric test. Is there a significant association between systolic blood pressure and BMI\_category? If yes, which BMI category has the highest and which BMI category has the lowest mean systolic blood pressure?
- d. Check that all assumptions of the parametric test performed are met. Provide all relevant screenshots.
- e. Based on the results of your assumption tests, conclude on whether the parametric test you performed in part b) above, is indeed an appropriate test to answer this research question. If it is not, please name an alternative test which could be used to answer this research question, and justify your answer (you do not need to perform the test; only to name it and justify your choice).

**Question 3**

Then, the researches wished to investigate whether systolic blood pressure before (“sysbp\_before”) and after (“sysbp\_after”) a healthy diet and physical activity intervention were significantly different in overweight individuals (BMI\_Category==2).

- a. Identify the independent/exposure variable and the dependent/outcome variable. State the type of each of these two variables where relevant.
- b. Identify the most appropriate bivariate test to test this relationship, justifying your choice of test in terms of the types of variables involved and the normality of any numeric variables. Run the test and provide the relevant screenshot.
- e. Interpret the results. Is there a significant difference in systolic blood pressure between before and after the intervention in overweight individuals? If yes, interpret the direction of the relationship / effect estimate.
- c. If the assumptions of the test performed in part b) above were not met, which statistical test could be used to investigate the difference in systolic blood pressure between before and after the intervention? You do not need to perform the test; only to name it and justify your choice.

**Question 4**

Lastly, the study group wanted to investigate whether baseline blood pressure (“*sysbp\_before*”) is associated with weight (“*weight*”) in the study participants. Therefore, they wished to investigate whether baseline blood pressure is significantly determined by the weight of participants.

- a. Identify the independent/exposure variable and the dependent/outcome variable. State the type of each of these two variables.
- b. Identify the most appropriate univariate test to test the magnitude of this relationship (by how much is systolic blood pressure influenced by changes in weight?), justifying your choice of test. Run the test and provide the relevant screenshot.
- c. Interpret the results. Is there a significant association between baseline systolic blood pressure and weight? If yes, interpret the coefficient for this association.
- d. Check that all assumptions of the test performed are met. Provide all relevant screenshots.

**Question 5 – Reflection**

Please reflect on your thinking, process of learning and work in this course up to date and also on the process of completing this assignment.

- i. Do you feel like you are learning/achieving the learning objectives of this course?
- ii. Are the strategies, skills and procedures you used to date effective for your learning?
- iii. What are some of the challenges that you faced but managed to overcome?
- iv. What can you do now to navigate the road ahead with the most success and overcome any existing challenges?
- v. What have you learned about yourself and about how you learn during this course?
- vi. What are you most proud of?

### 3. Submission, assessment and feedback

- The Report (Word document) will have to be submitted via Moodle.
- Deadline for submission: **Friday, 5<sup>th</sup> of May 2023, 11:59 pm**
- A Marksheet with specific marking criteria will be used to assess each student separately (Annex 1 at the end of this document).
- The total grade of the assignment is 100 marks.
- Written feedback will be provided.

The report contributes towards 30% of the overall grade of the MPH-512 course.

**ANNEX 1: MPH 512 Summative Assignment**

**Spring 2023**

**MARKSHEET**

<b>Course:</b>	MPH-512: Applied Statistics for Epidemiology and Public Health	
<b>Course Lead:</b>	Dr Christiana Demetriou	
<b>Student Name and Student ID:</b>		

Assignment 1 - Mark scheme

QUESTIONS	ALLOCATED MARKS	STUDENT MARKS
<p><b>Question 1</b></p> <p>Researchers wished to investigate the smoking status (“smoker”), as a categorical predictor variable, with respect to its association with being overweight or obese:</p> <p>a. <u>Generate</u> the variable “BMI_category” and <u>tabulate</u> it.</p> <p>b. Create labels for the BMI_category variable. Attach the labels to the variable. Re-tabulate the variable.</p> <p>c. Run the most appropriate univariate test to test this relationship and provide the relevant screenshot. Justify your choice of test.</p> <p>d. Interpret the results.</p>	<p>6</p> <p>4</p> <p>10</p> <p>2</p>	

<p><b>Question 2</b></p>	
<p>Following, the study group wanted to investigate whether BMI_category is associated with systolic blood pressure (“sysbp_before”) in the study participants.</p>	
<p>a. Identify the independent/exposure variable and the dependent/outcome variable. State the type of each of these two variables</p>	<p>4</p>
<p>b. Identify a <u>parametric</u> univariate test to test this relationship, justifying your choice of test in terms of the variables involved. Run the test and provide the relevant screenshot.</p>	<p>10</p>
<p>c. Interpret the results.</p>	<p>4</p>
<p>d. Check that all assumptions of the parametric test performed are met. Provide all relevant screenshots.</p>	<p>12</p>
<p>e. Identification of an appropriate alternative test (if relevant)</p>	<p>2</p>

<p><b>Question 3</b></p>	
<p>Then, the researches wished to investigate whether systolic blood pressure before (“sysbp_before”) and after (“sysbp_after”) a healthy diet and physical activity intervention were significantly different in overweight individuals (BMI_Category==2).</p>	
<p>a. Identify the independent/exposure variable and the dependent/outcome variable. State the type of each of these two variables where relevant.</p>	<p>2</p>
<p>b. Run the most appropriate univariate test to test this relationship and provide the relevant screenshot. Justify your choice of test.</p>	<p>10</p>
<p>c. Interpret the results.</p>	<p>4</p>
<p>d. If the assumptions of the test performed in part b) above were not met, which statistical test could be used to investigate the difference in systolic blood pressure between before and after the intervention?</p>	<p>2</p>

<b>Question 4</b>			
<p>Lastly, the study group wanted to investigate whether baseline blood pressure (“sysbp_before”) is associated with weight (“weight”) in the study participants. Therefore, they wished to investigate whether baseline blood pressure is significantly determined by the weight of participants.</p>			
<p>a. Identify the independent/exposure variable and the dependent/outcome variable. State the type of each of these two variables</p>		2	
<p>b. Run the most appropriate univariate test to test the magnitude of this relationship (i.e. by how much is systolic blood pressure influenced by changes in weight?) and provide the relevant screenshot. Justify your choice of test.</p>		10	
<p>c. Interpret the results. Is there a significant association between weight and hypertension?</p>		4	
<p>d. Check that all assumptions of the test performed are met. Provide all relevant screenshots.</p>		12	
<b>Total Marks</b>		100	0

**General Comments:**

**STRENGTHS**

**AREAS FOR IMPROVEMENT**